# Catalogue de formations 2023Sommaire

Kiran Derennes | Ifig Le Gonidec | Clémence Legendre | Mattéo Rouanne

SAE 501, Analyse et Conception d’un outil décisionnel

16/01/2025

[Introduction/Contexte](#_Toc188428881)

[Traitement des données](#_Toc188428882)

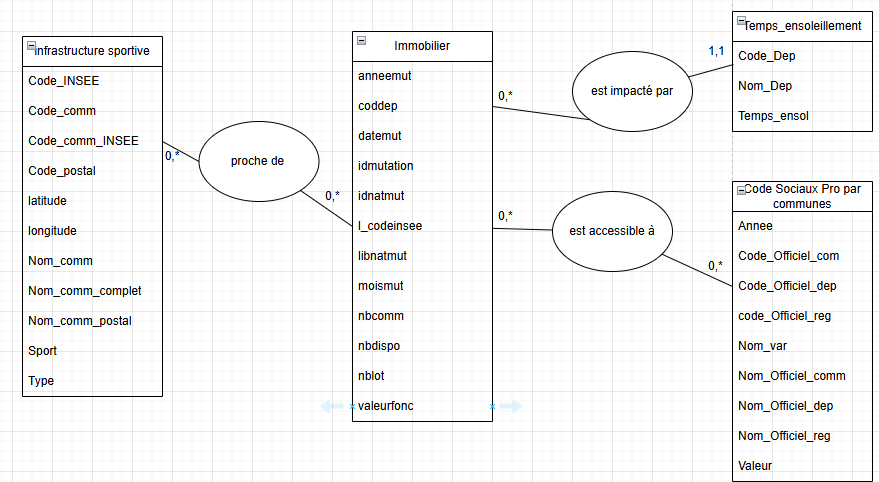
[Integration des données](#_Toc188428883)

[Conclusion](#_Toc188428884)

# Introduction/Contexte

En 2024, la France comptait près de 36 millions de logements. Cette vaste offre de logements reflète les défis qu'ils représentent dans un contexte démographique et économique tendu. Pourtant, les demande et les offres restent inégales selon les régions, avec des tensions particulièrement marquées dans les grandes métropoles et les zones touristiques, où les prix ont connu des hausses considérables ces dernières années. Dans la recherche d’un logement que nous soyons en couple, célibataire, personne âgée ou un étudiant, il y a de nombreux critères qui nous feront prendre un logement ou un autre. Nous allons à travers ce projet décrire la situation actuelle de l’immobilier en France en croisant les logements achetés avec d’autres informations afin d’expliquer le choix du logement. Nous vous voulons donc répondre à cette problématique : Qu’es ce qui justifie l’écart de valeur foncière entre les Pays de la Loire et la Bretagne et entre leurs départements ? Pour répondre à cette question nous avons donc étudié les données qui pourrait aider à répondre à la question, mais aussi comment les présenter à un client qui serait intéresser de connaître la réponse à travers un système décisionnel.

Pour expliquer cette décision, nous avons tout d’abord décider de nous concentrer uniquement entre 2 régions, la Bretagne et Pays de la Loire. Nous avons donc pris les données sur les logements qui nous intéressent. Nous avons aussi à travers les données proposées par l’INSEE et OpenDataSoft, choisis les données sur les communes, département et régions et les temps d’ensoleillement par jour par an dans le département. Les codes sociaux professionnels par communes pourraient expliquer la décision de même pour les infrastructures sportives proposé par commune, département et région.



Modèle Relationnel 1- Modèle relationnel des données

Voici notre modèle relationnel de notre jeu de données sur lequel nous allons travailler afin de préparer un outil décisionnel. Nous retrouvons toutes les tables de données mentionnées précédemment mais nous ajoutons ici les variables que nous utiliserons. Nous avons trié au préalable afin de ne garder que les variables qui pourraient nous aider à répondre à la problématique. Nous pouvons aussi observer les relations entre les dimensions du modèle.

Nous allons préparer les données sur logiciel ETL Pentaho, où nous allons nettoyer et homogénéiser les données de manière automatique afin qu’il ne suffise que de mettre à jour les données pour avoir les résultats d’analyses actualisé. Puis nous analyserons les données et tenterons de répondre à la problématique à travers un outil d’aide à la décision Power BI.

# Traitement des données

Nous avons préparé les données sur le logiciel Pentaho afin que les jeux de données soient utilisables entre eux sur notre outil d’analyse Power BI. Notre dictionnaire des données[[1]](#footnote-1)consiste en 4 tables décrivant les informations sur l’immobilier en région Pays de la Loire et de la Bretagne ainsi que les infrastructures disponibles dans ses régions les codes sociaux professionnels établis dans les régions et enfin les temps d’ensoleillement en jours par an dans les régions définit.

Une image contenant texte, diagramme, ligne, capture d’écran

Description générée automatiquement

Pentaho - nettoyage des données immobilier

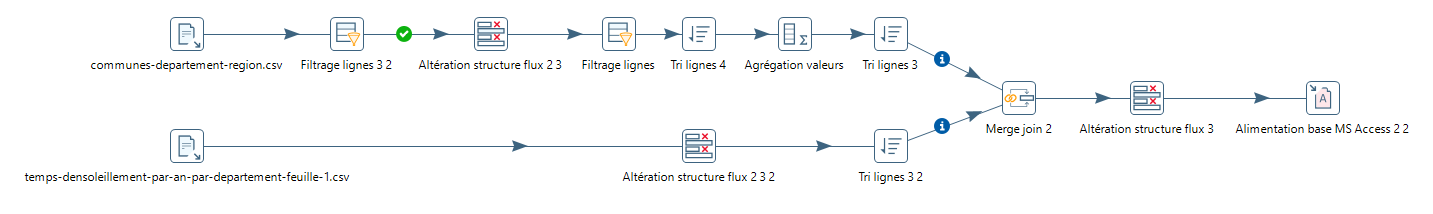
Dans un premier temps, nous avons importé les données immobilières pour tous les départements des régions Bretagne et Pays de la Loire. Celles-ci, proviennent du site Cerema. Nous avons fusionné les jeux de données afin de n’en faire ressortir un seul fichier pour que ça soit plus simple d’analyser et comparer les régions. Ensuite nous avons recoder la variable du Code INSEE afin de retirer les crochets qui entouraient les codes. Nous avons filtré les données immobilières pour ne garder uniquement celles qui concernes les ventes et nous avons supprimer les variables qui ne nous intéressaient pas. Le modèle relationnel décrit les variables que nous avons gardé. Nous avons enfin exporté le jeu de données dans un base de données Access, dans une table crée à cet effet : table\_des\_faits.

Une image contenant ligne, diagramme, Tracé, Police

Description générée automatiquement

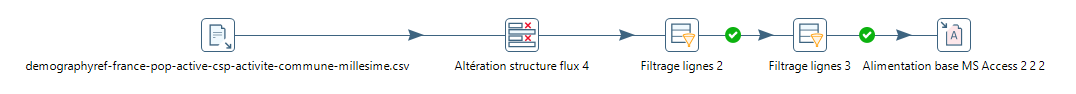
Pentaho - création du fichier structures sportive

Nous avons ensuite importé la table sur les infrastructures sportives, oms-France-sport-facility, ainsi que la table sur les communes, département et régions. Ce fichier a été trouvé sur OpenDataSoft et nous avons chacun vérifier pour nos communes respectives la véracité du jeu de données qui nous semblaient tout à fait correct. Pour cette table, nous avons juste supprimer les variables qui ne nous servaient pas alors que pour la table des communes nous avons filtrer la table sur les régions qui nous concernait ainsi lors de la jointure des tables la tables des infrastructures sera filtrer sur les mêmes régions car on les joints sur les codes de départements. Enfin nous avons exporté la table dans une base de données Access, dans une table crée à cet effet : table\_ communes.



Pentaho - création du fichier temps d'ensoleillement

Pour la table du temps d’ensoleillement, nous avons importé le jeu de données venant du site de l’INSEE associé à cette table puis nous avons supprimé les variables qui ne nous intéressaient pas. Mais nous avons aussi réutiliser le jeu de données communes département et région où nous avons filtrer sur les régions nous concernant et supprimer les variables inutiles. Nous avons aussi fait attention à ne pas garder les données dont le code département était nulle. Nous avons dû ajouter une étape pour regrouper par département car dans le jeu de données certains département était dédoublonné. Puis nous avons joint les tables sur le temps d’ensoleillement et des communes dans lequel nous avons supprimé les dernières variables superflues comme les variables qui existaient dans la table commune et temps d’ensoleillement avant d’extraire la jointure des tables dans une base de données Access, dans une table crée à cet effet : table\_ departements.



Pentaho - création de la table CSP

Nous avons importé le jeu de données sur les codes sociaux professionnels, comme pour les autres jeux de données nous avons supprimer les variables superflues. Le jeu de données CSP prend en compte une certaine plage d’information aussi nous avons fait le choix de nous concentrer sur l’année 2019 illustré par le filtre sur les années ainsi sur les régions Bretagne et Pays de la Loire à partir d’un autre filtre. Nous avons ensuite exporté la table dans une base de données Access.

Nous avons utilisé la table Communes département et région afin de préparer les jointures entre les tables créer sur Pentaho néanmoins nous avons estimer que la table en elle-même n’étaient pas intéressant pour l’analyse des résultats sur notre outil PowerBI.

Afin de rendre l’extraction et le nettoyage des données plus automatique, nous avons préparer un job qui lance les programmes de nettoyage pour chaque table d’un seul click au lieu de le faire manuellement pour chaque table.

Une image contenant Police, texte, diagramme, ligne

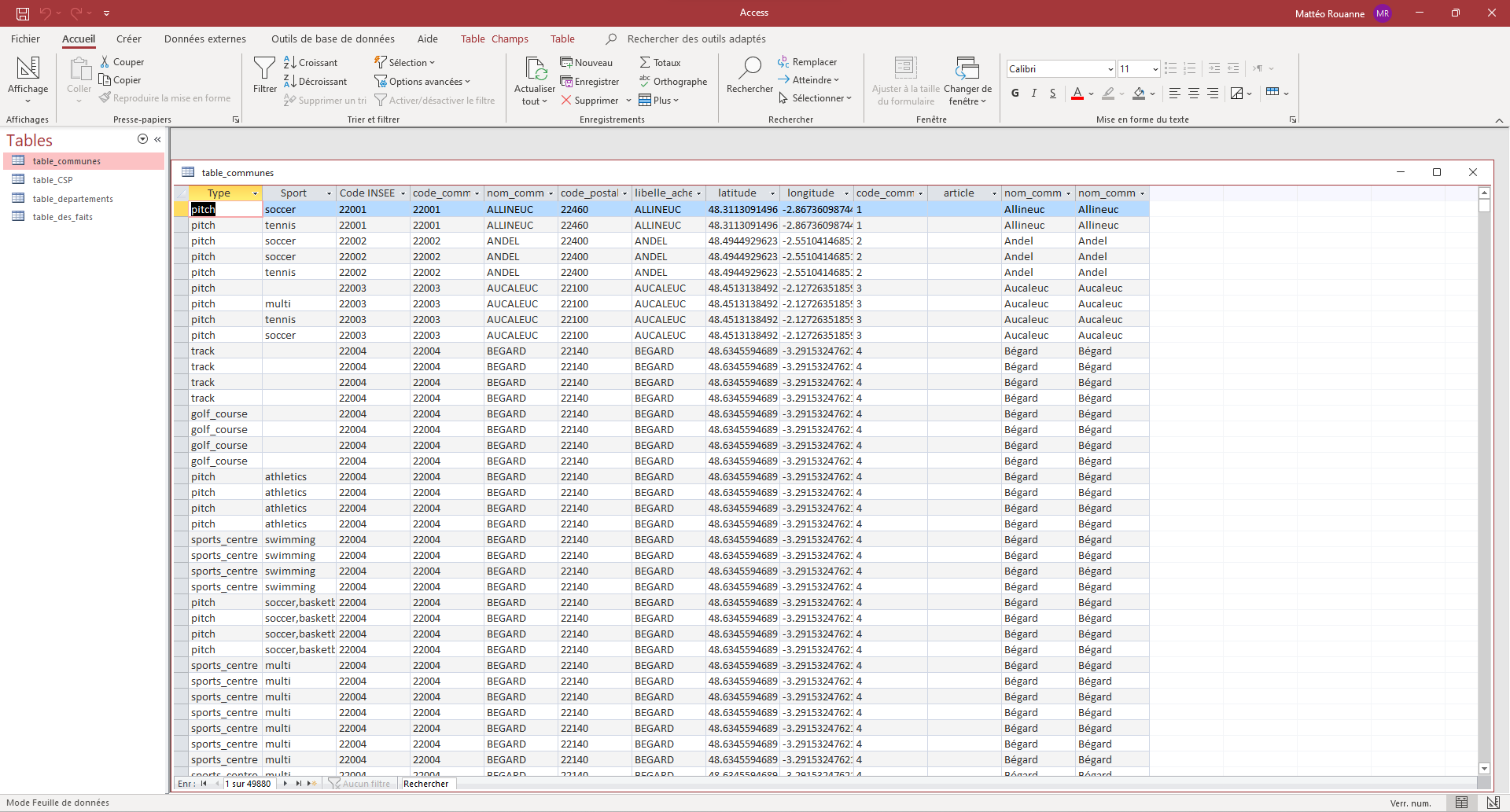
Description générée automatiquement

Pentaho 5 – Job de lancement des programmes de nettoyage

Ce programme consiste en un Start qui permet quand lancer d’exécuter chaque programme un à un. Chaque exécution fait référence au programme créer dans notre job pour importer les jeux de données nécessaires, nettoyer le jeu de donnée et l’extraire dans une base de données Access. Si nous voulions aussi automatiser ces programmes pour qu’ils soient lancer de manière régulière nous pourrions le faire à partir de ce job.

# Intégration des données

Nous avons créé une base donnée Access pour chaque table créée dans Pentaho afin de pouvoir rendre l’intégration des données automatisé. Ainsi sur Access nous retrouver les 4 tables crées précédemment qui sont par la suite connecté à notre outil Power BI.



Base\_Access - Intégration des tables créés sur Pentaho

Ces tables pourront donc être mise à jour régulièrement afin que notre outil Power Bi puisse présenter des informations actualisées. Nous avons rencontré quelques difficultés pour intégrer nos tables dans notre base de données Access car dans un premier temps les programmes afin d’exporter les données étaient dans un même job. Au début, nous n’exportions pas les données sur des bases mais sur des fichier CSV et quand nous les exportions il n’y avait pas de problèmes. Mais dès lors où nous avons décider de les exporter sur des bases de données, à chaque fois que nous exécutions chaque programmes les tables se superposais et se remplaçait par la table exporter ensuite. Il a donc fallu créer un job par extraction de tables et les actionner à partir du job d’exactions afin que les tables se retrouvent correctement dans les bases de données sur Access.

# Conclusion

Tout le travail précédemment décris nous a permis de concevoir un outil d’analyse sur Power BI qui nous permet de mieux répondre à notre problématique.

Une image contenant texte, carte, capture d’écran, Police

Description générée automatiquement

Depuis notre outil Power BI, on remarque une hausse générale des prix. En effet, les valeurs foncières ont connu une augmentation significative sur la période étudiée. À cela s’ajoute des disparités territoriales marquées, les prix varient considérablement selon les régions et les départements, reflétant des dynamiques locales spécifiques. Notamment entre la Mayenne et la Loire-Atlantique où les valeurs foncières de la Mayenne sont deux fois moins importantes comparer à la Loire-Atlantique. On peut également remarquer l’influence des zones urbaines, les zones à forte densité urbaine et les secteurs en développement ont généralement connu des hausses de prix plus importantes.

Une image contenant texte, capture d’écran, logiciel, Système d’exploitation

Description générée automatiquement

Concernant le croisement de nos données avec les catégories socio-professionnel on constate une dominance des catégories socio-professionnelles « supérieures ». En effet, les cadres et les professions intermédiaires représentent une part importante des acheteurs, témoignant d'un pouvoir d'achat plus élevé.

Une image contenant texte, capture d’écran, Police, nombre

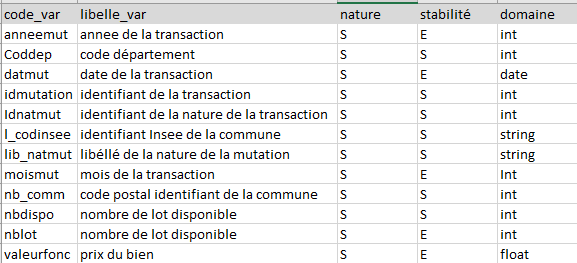
Description générée automatiquement

Finalement, on constate une croissance soutenue des prix, les deux régions ont connu une hausse significative des prix immobiliers sur la période étudiée. Toutes deux suivent une évolution quasi similaire hormis pour l’année 2019 ou la moyenne des valeurs foncières en Bretagne a explosé. Ceci est dû au développement du système ferroviaire breton, à la forte tendance d’achat de maisons secondaires dans les régions côtières et au taux de prêt très bas à l’époque.

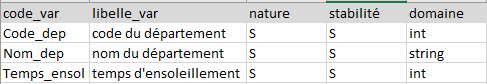
En conclusion, nous pouvons affirmer qu’il y a :

* **Un marché immobilier dynamique :** Le marché immobilier français a connu une croissance soutenue sur la période étudiée, portée par une demande soutenue et une offre limitée dans certaines zones.
* **Des disparités territoriales persistantes :** Les prix et les volumes de transactions varient considérablement selon les régions, reflétant des contextes économiques, démographiques et sociologiques différents.
* **Une évolution des profils d’acheteurs :** Les catégories socio-professionnelles supérieures jouent un rôle de plus en plus important sur le marché.

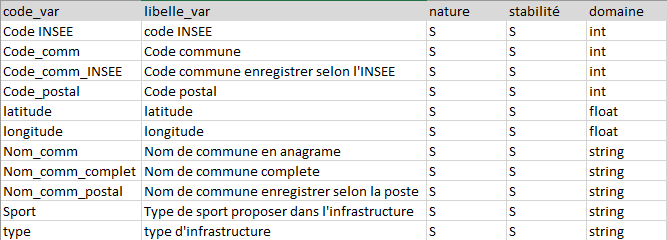
# Annexe

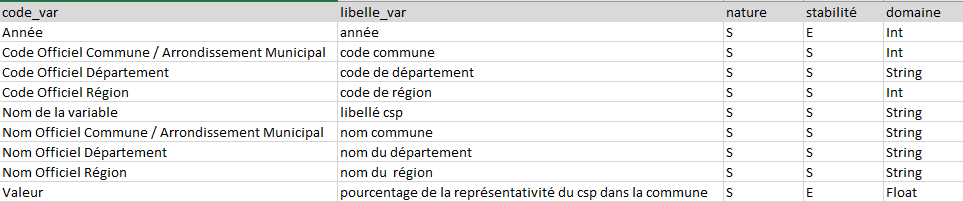


Dictionnaire des Données - Table Immobilier



Dictionnaire des Données 2 – Table Temps\_ensoleillement

Dictionnaire des Données 3- Table infrastructure sportive

Dictionnaire des Données 4 - Table CSP par communes

1. Le dictionnaire des données pour chaque table est accessible en annexe. [↑](#footnote-ref-1)